

Harsh Patel

USA (Open to relocate) | +1 (315) 928 8459 | harshp1517@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Software Engineer with 3+ years of experience designing and shipping scalable backend services, distributed microservices, and LLM-powered intelligence systems across marketplace and enterprise SaaS platforms. Owns the full engineering lifecycle from API design and database modelling through cloud deployment and production observability. Proficient in Python, FastAPI, Java, Spring Boot, React, PostgreSQL, Redis, and AWS (Lambda, ECS, SQS, API Gateway, DynamoDB), with hands-on experience integrating RAG pipelines, vector search, and LLM classification into high-throughput production systems. Consistent track record of improving system reliability, reducing p95 latency, and eliminating manual operational overhead through clean architecture, async processing patterns, and robust CI/CD automation.

SKILLS

Languages & Frameworks: Python, Java, JavaScript, TypeScript, SQL, Shell Scripting, FastAPI, Spring Boot, Spring MVC, Spring Batch, Hibernate, React, Redux

AI Engineering: LLM APIs, LangChain, Retrieval-Augmented Generation, Text Embeddings, Vector Databases (pgvector, FAISS), Prompt Engineering, Agentic Workflows, Structured Output Validation, AI Guardrails, LLM Evals, NLP Pipelines, Hugging Face, GitHub Copilot

Databases: PostgreSQL, DynamoDB, MongoDB, MySQL, Redis, query optimization, connection pooling, schema design

Cloud & DevOps: AWS (Lambda, ECS, API Gateway, IAM, CloudWatch, SQS, S3), Docker, GitHub Actions, Jenkins, CI/CD Automation, Containerized Deployments

System Design & Architecture: Distributed Systems, Microservices, Event-Driven Architecture, Load Balancing, Rate Limiting, Circuit Breaker Patterns, SLO/SLI Definitions, gRPC, WebSockets

EXPERIENCE

Honeywell

Software Developer (Co-op)

Sep 2025 - Present

- Architected a seller-facing operations platform using React, Node.js, and DynamoDB with real time slot enforcement, conflict detection, and role-based access control, enabling 80% of sellers to self-manage daily workflows and cutting ops intervention requests by 45%.
- Shipped low-latency pricing and availability microservices in Python and Node.js with async event handlers, time-window validation, and idempotent write patterns, reducing booking conflicts by 33% and sustaining sub-150ms p95 response times under peak load.
- Designed an LLM-powered onboarding pipeline using Python/FastAPI, LangChain, and OpenAI with prompt engineering, structured output parsing, and JSON schema validation, converting unstructured seller inputs into standardized config rules and cutting clarification cycles by 40%.
- Built a RAG-based semantic search service using LangChain, text embeddings, and pgvector with approximate nearest-neighbor indexing, surfacing contract clauses and setup guidance in under 2 seconds and reducing first week support tickets by 25%.
- Streamlined CI/CD release pipelines using GitHub Actions, Docker, and AWS ECS with blue green deployments, automated smoke tests and staged rollouts, cutting deployment time from 30 to under 10 minutes and eliminating environment-parity failures across staging and production.

Mindtree

Software Developer

Jan 2022 - Jul 2024

- Owned end-to-end design and delivery of Java Spring Boot microservices for a B2B SaaS workflow platform, covering approval routing, role-based task assignment, and audit-trail persistence in PostgreSQL, serving 800+ enterprise clients and 50K+ daily active users with sub-200ms p95 response times under concurrent load.
- Implemented RESTful APIs using Spring MVC and Hibernate with layered validation, optimistic locking, and indexed query strategies in PostgreSQL, cutting p95 API latency from 1.8s to 320ms and supporting 3x user growth without infrastructure scaling.
- Established async job processing with Spring Batch and Redis for scheduled data syncs, notification dispatch, and cleanup workflows using idempotency keys, retry policies, and exponential backlog, eliminating job-overlap failures and improving pipeline reliability by 30%.
- Delivered React-based dashboard UI integrating Spring Boot REST endpoints, featuring drag and drop Kanban boards and role-based views, improving enterprise client engagement by 35% and reducing onboarding time by 40%.
- Containerized all Spring Boot services using Docker with health checks and environment scoped configs and automated deployments via CID pipelines, reducing post-release defects by 37% across multiple delivery cycles.

EDUCATION

Master of Science in Computer Science

May 2026

Syracuse University

Syracuse, NY

PROJECT

Space Application Centre, ISRO, India.

- Deployed ML-driven models using Python, Scikit-learn, and PyTorch, to successfully predict the geo-climatic conditions over the Indian subcontinent by an extent.
- Studied and gathered Satellite Imagery, Radar and LiDAR data, Radiometric data from the INSAT 3D satellite and the DPR radar to build a pipeline for the database by normalizing and flattening the data into a reusable database.

CERTIFICATION

- AWS Solution Architect, AWS Academy Machine Learning Foundations
- Azure AI Fundamentals
- SAP Certified Application Associate - SAP Analytics Cloud
- Google Cloud Platform Badges